# The Small Angle Scattering Toolbox

**An open source software framework for the analyses of bioSAXS data**

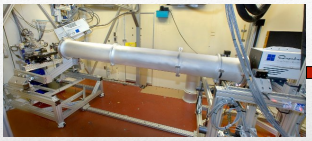
Peter Zwart

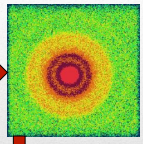
Lawrence Berkeley National Laboratories

http://sastbx.als.lbl.gov

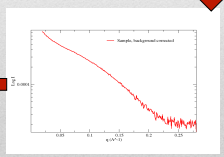
PHZwart@lbl.gov

# Introduction





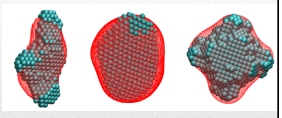
- Information:
  - Size
  - Weight
  - Shape
  - ….

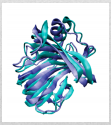


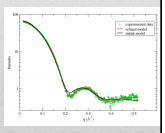
- Advantages
  - Near native condition
  - Dynamic
- Disadvantage
  - Low information content

# The SASTBX

- The SASTBX: open-source platform for small angle scattering analyses
  - http://sastbx.als.lbl.gov
- Extends the CCTBX
- Functionality
  - Data reduction
  - Model data calculation
  - P(r) fitting
  - Guinier/Kratky analyses
  - **Shape retrieval**
    - No intensity preprocessing required
  - **Structure refinement**
    - Normal mode based, against intensity
  - **And beyond**



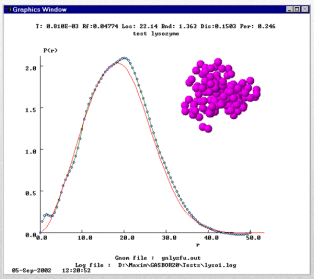




# From a curve to a shape in seconds

- Fast database searches have transformed biological sciences
  - Google (ipad pricing)
  - BLAST (sequence searching)
  - SSM (Structure searching)
- Given a SAXS curve, we would like to have a rough idea of the associated shape within seconds of availability of the data: *sastbx.shapeup*
  - name suggested by Jack Tanner, U Missouri
- Convert difficult inverse problem into straightforward optimization problem

## Traditional methods



- Optimize real space model with respect to data and spatial compactness restraints via Simulated Annealing techniques
  - Dummy atoms
  - Slow

## sastbx.shapeup

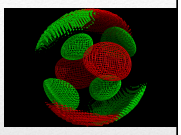
- Real space 3D models parameterized by 3D Zernike polynomials
  - Fast intensity calculation (order of magnitude)
  - Database of shapes (PDB, PISA, PIQSI)
    - Have a large set of biologically inspired shapes
    - Precompute things: need for speed
- Shape *reconstruction* transformed into shape *retrieval*
  - Two parameter optimization problem
    - Size
    - Location in the database / shape index

## 3D Zernike polynomials

- 3D Zernike polynomials

$$Z_{nlm}(\boldsymbol{r}) = R_{nl}(r)Y_{lm}(\theta,\phi) \;\; with \;\; \boldsymbol{r} = (r,\theta,\phi)$$

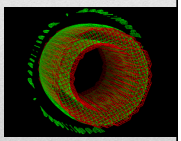
  - Orthonormal in unit sphere



- 3D Zernike moments
  - Can be computed fast using linear combination of geometric moments

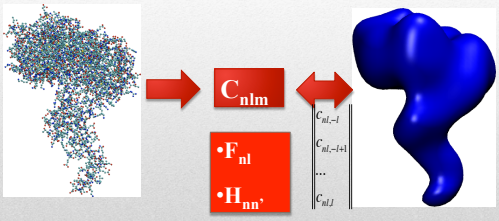
$$c_{nlm} = \frac{3}{4\pi}\int_{|\boldsymbol{r}|<1} \rho(\boldsymbol{r}) Z^{*}_{nlm} d\boldsymbol{r}$$

  - Series of Zernike polynomials can be used to build any shape in the 3D ball



$$\widehat{\rho}(\boldsymbol{r}) = \sum_{n=0}^{n_{max}} \sum_{l=0}^{n} \sum_{m=-l}^{m=l} c_{nlm} Z_{nlm}(\boldsymbol{r})$$

## Model representation

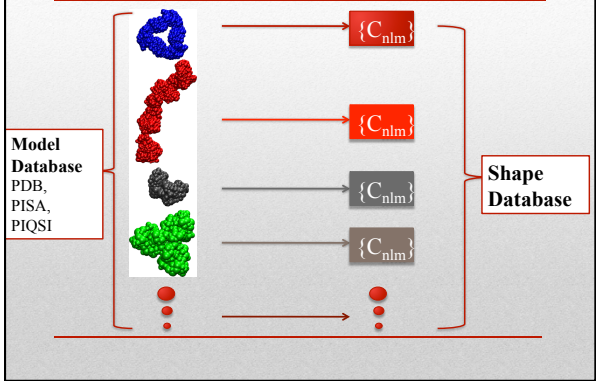



{(x,y,z)} 1000x

{$C_{nlm}$} 100x

* Mak & Morris (2008)

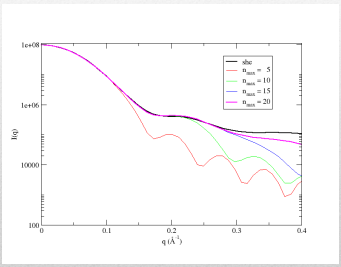
## Shape database




## Intensities from 3DZM

$$I(q) = \sum_n \sum_{n'} B_{nn'}(qr_{\max}) B_{n'n'}(qr_{\max}) H_{nn'}$$

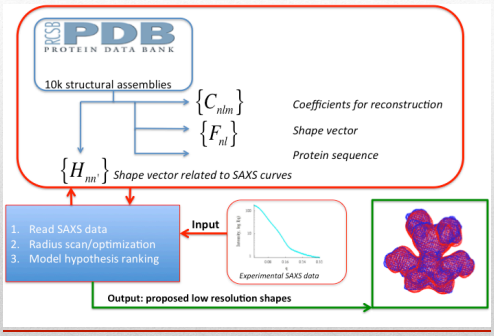
$$H_{nn'} = \sum_l k_{nl} k_{n'l} \sum_m c_{nlm} c^*_{n'lm}$$




SAXS intensity is a linear combination of 3D Zernike moments, weighted by products of spherical Bessel functions

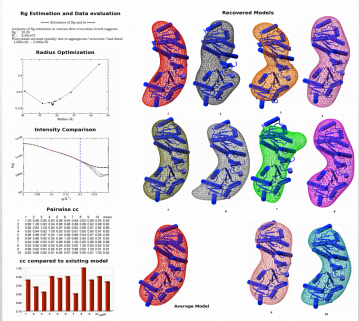
Particle size rmax 'stretches' the curves

## Shapeup procedure



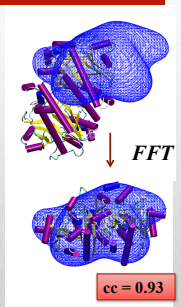

## sastbx.shapeup data=data.qis

- Get a coffee
  - But make it quick: you have 80 seconds



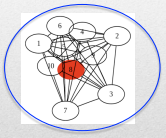
## Aligning Zernike Models

- Coefficients of rotated real space Zernike models can be easily computed
  - rotating spherical harmonics, Wigner D matrices
- The correlation coefficient for all rotations can be computed using a single 3D FFT.
  - Initial hits can be optimized by a local simplex based search
- This is in essence the same as a crystallographic rotation function

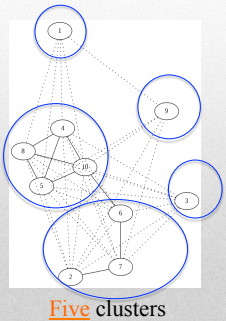



Trapani and Navaza (2006)
Kovacs and Wriggers(2002)

## Internal consistency



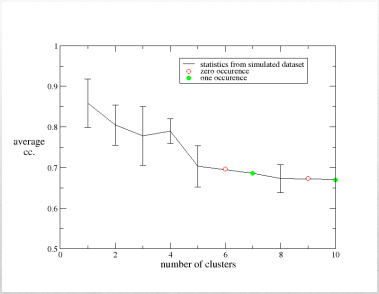
One cluster



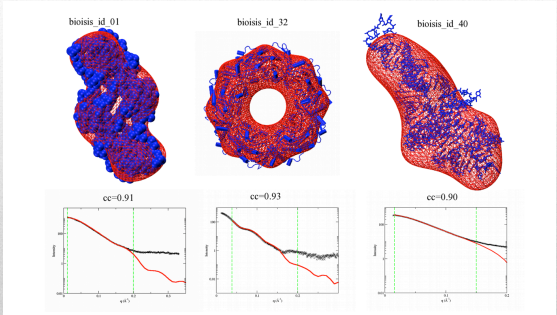
Five clusters

## A good solution is a consistent one

- Model Quality is positively correlated to the consistency
  - Number of model clusters =1: very consistent models
  - Number of model clusters =2: okay
  - Number of model clusters >2: ??? (not good)

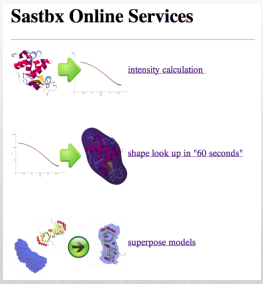



## Experimental data



# Online interface

- The most straightforward manner is to run the shapeup using our online server
  - But you can install a local copy if desired
- http://sastbx.als.lbl.gov
  - Go to online services
  - Select shape lookup
  - Upload your SAXS data / reference model
  - Wait ~60 seconds
  - View maps online

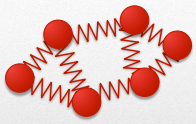



# Structure refinement

- Small Angle Scattering provides powerful information of the shape of macromolecules in solution
- Although we can obtain some information from a low resolution shape, having an atomic model (or ensemble of atomic models) is often more informative
- In a lot of cases we already have a good idea how the molecule looks
  - NMR
  - Crystallography
- The X-ray structure however often doesn't fit the SAXS curves in a satisfactory manner
  - We need refinement

# Structure refinement

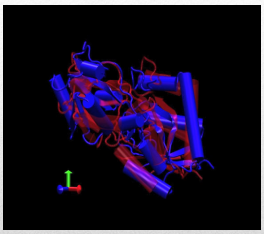
- Refinement?
  - Take structure
  - Perturb it (how?)
  - Compute SAXS/WAXS data
  - Check if agreement between calculated and observed data improves
- How to perturb?
  - Normal modes
  - Random
  - Gradient based
- Outstanding issues
  - Radius of convergence
  - Solvent

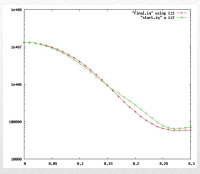


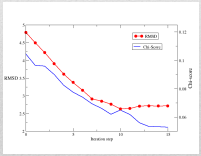
Proteins represented by an elastic network allow low-energy (i.e. likely) deformations to be computed using eigenvalue analyses

1. Compute normal modes
2. Select normal modes
3. Do local simplex optimization by normal mode perturbations
4. Fix geometry (PULCHRA)
5. Go to 1

# Refinement







LAO Binding Protein

RMSD: 4.8Å → 2.2 Å

Synthetic data

## Discussion

- Quoted timings are on a single processor
- Most time is spend on building and aligning maps
  - Reading in the database (>10k entries) takes less then 2 seconds
  - The shape retrieval is done within 20 seconds
  - No symmetry is currently imposed
- Target function is not a simple Chi-square target but a 'likelihood inspired, empirically derived function that seems to fit the bill but has to be improved upon'
  - Better then basic Chi Squares
  - Resolution dependent error term for a 'fixed' model error and estimated particle size

## Thank you

Ralf Grosse-Kunstleve
Paul Adams

Alex Hexemer
Eric Schaible
Robert Rambo

Jack Tanner

Funding from LBNL

http://sastbx.als.lbl.gov